

How Voting and Consensus Created the Diagnostic and Statistical Manual of Mental Disorders (DSM- III)

James Davies

Anthropology and Medicine

To cite this article: James Davies (2016): How Voting and Consensus Created the Diagnostic and Statistical Manual of Mental Disorders (DSM-III), Anthropology & Medicine, DOI: 10.1080/13648470.2016.1226684

Abstract:

This article examines how Task Force votes were central to the development of DSM-III and DSM-III-R. Data were obtained through a literature review, investigation of DSM archival material housed at the American Psychiatric Association (APA), and interviews with key Task Force members of DSM-III and DSM-III-R. Such data indicate that Task Force votes played a central role in the making of DSM-III, from establishing diagnostic criteria and diagnostic definitions to settling questions about the inclusion or removal of diagnostic categories. The article concludes that while the American Psychiatric Association represented DSM-III, and the return to descriptive psychiatry it inaugurated, as a triumph of empirically-based decision making, the evidence presented here fails to support that view. Since the DSM is a cumulative project, this article calls for a more socio-historically informed understanding of DSM's construction to be deployed in how the DSM is taught and implemented in training and clinical settings.

Keywords: Borderline Personality Disorder; Diagnostic and Statistical Manual of Mental Disorders (DSM); Diagnosis; Robert Spitzer; Self-Defeating Personality Disorder.

The publication of DSM-III in 1980 was a response to diverse social-cultural and clinical pressures bearing on psychiatry in the late 1960s and early 1970s. Research by Rosenhan (1973) Cooper, et al., (1972) and Gurland (1972) had exposed the problem of low diagnostic reliability, giving impetus to those critics already arguing that psychiatric diagnosis was wrongly medicalizing many problems of living. Poor reliability also threatened psychiatry's perceived capacity to provide reliable diagnostic data for the facilitation of psychotropic research; an issue heightened by the Thalidomide crisis in the early 1960s, which obliged the FDA to accept the need for randomized, placebo-controlled trials to prove drug safety and effectiveness (Junod 2008). In addition, some insurers dissatisfied with the imprecision of DSM-II, now wanted a clearer and coded diagnostic system for reimbursement purposes; one that would also ideally better cohere with the codes in ICD-9. Finally, the growing number of non-medical mental health providers during the 1970s threatened psychiatry's status and relevance, especially since low diagnostic reliability potentially discredited one of the linchpins of psychiatric research and practice. DSM-III was therefore launched with the aim of creating an a-theoretical, empirically-based diagnostic system that would address the problems ensuing from the poor reliability besetting previous DSM editions.

From the standpoint of the American Psychiatric Association (APA) the manual that ultimately emerged was a success. DSM-III's scientific aura satisfied the regulators,

the insurers, the drug companies and also re-established what psychiatry could uniquely offer (Tietze 2014). From a scientific and clinical standpoint, however, many critics argued that DSM-III created as many problems as it purportedly solved, not for the profession but for prospective patients as it exacerbated the unnecessary medicalizing of much socially induced suffering, which in turn could lead to inappropriate and extensive prescribing of pharmaceuticals (Conrad 2007, 1992, Kutchins and Kirk 1998, Szasz 2007). They argued the text was a triumph of rhetoric not science, as the process of its construction was murky and its methods marred by too many subjective variables (Kirk and Kutchins 1992). It also, most importantly, failed to deliver on its *raison d'être* - solving the reliability problem – since its field trials revealed many of its categories to be hardly more reliable than those of DSM-II (Kutchins and Kirk 1986, 1994, Eysenck 1986, Scheff 1986).

The extensive and now well-documented criticisms of DSM-III raise the question as to why, 35 years after its publication, people are still curious and critical regarding its construction. A number of factors can help explain this. The recent controversy over DSM-5 has renewed interest in the epistemological and methodological underpinnings of the broader DSM project, which are rooted to a large extent in DSM-III. In addition, as DSM-III broadly established the modern diagnostic system under which users of subsequent editions still largely operate, and as many of its diagnostic categories and criteria sets have significantly influenced those of DSM-IV and DSM-5, each edition builds cumulatively on the last making analysis of previous editions integral to any analysis of the total project. A further reason for the enduring interest in DSM-III is that much of the archival material regarding its construction still remains unread, unanalyzed and unpublished. Admittedly, some useful and

comprehensive archival-based studies have emerged since its publication, the most recent and notable being Hannah Decker's (2103) *The Making of DSM-III*, which diligently charts the political maneuverings and controversies besetting DSM-III's construction. Archival evidence also appears in historical and philosophical studies such as Edward Shorter's (2009) illumination of the unscientific broadening of major depressive disorder, Mitchell Wilson's (1993) analysis of how DSM narrowed psychiatry's clinical gaze, Christopher Lane's (2007) work on how shyness was reclassified as social anxiety disorder, and Rachel Cooper's (2004) work on the extent to which DSM is 'theory laden'.

By and large, however, the extent and breath of the archival material (which extends, after all, to approximately 9 lineal feet) has ensured that much material of critical significance has remained un-sourced and unexamined. This is certainly true with respect to material pertaining to the extent to which Task Force votes were a primary mechanism by which deliberations and conclusions were settled on the Task Force, and in turn how the manual was constructed. This article will therefore illuminate this relatively neglected area through data obtained from firstly, previously unpublished archival materials sourced at the APA archives in Arlington VA., and secondly, from data gathered through interviews with members of, and consultants to, the DSM-III and III-R Task Forces.

Composition of the Task Force for DSM-III:

Work officially commenced on DSM-III in 1974. Its original Task Force, ratified by the American Psychiatric Association's Council of Research and Development, relied exclusively on the Task Force's Chairman, Robert Spitzer, to select and appoint its

members (Decker 2013: 148). Initially the Task Force comprised eight members whom Spitzer chose because of their “special interest in various aspects of diagnosis” (DSM 1980: 2). These members were: Nancy C. Andreasen (psychiatrist); Robert Spitzer (psychiatrist); Theodore Millon (psychologist); Donald F. Klein (psychiatrist); Henry Pinsker (psychiatrist); Jean Endicott (psychologist); George Saslow (psychiatrist - deceased); Robert A. Woodruff (psychiatrist - deceased); Morton Kramer (biometrician - deceased). The text editor was Janet B. W. Williams.

The Task Force’s purpose was to decide all matters related to how the manual would be researched and compiled. It would evaluate all proposals for DSM-III stemming from advisory committees, liaison committees, professional organizations within the APA and participants in the DSM-III Field Trials. The Task Force also aimed to represent the profession’s diversity, which led to its expansion as work progressed. As the introduction to DSM-III states, “additional members were added to ensure representation of different perspectives and areas of expertise” (DSM 1980:2). Indeed, at its height the Task Force comprised 15 members largely from various sub-fields of psychiatry. Whether this expansion sufficiently represented the diversity of views within psychiatry and the broader mental health professions remained moot. Early on the Chair of the APA’s council on R and D expressed to Spitzer concern about the absence of “minority members”, a concern also articulated by the Committee of Black Psychiatrists, who had no Task Force representation (Decker 2013: 148). Psychoanalytic presence was also kept to a minimum, which later generated protest from the American Psychoanalytic Association. In short, the Task Force essentially comprised white, middle-class Americans all, bar one, from the disciplines of psychiatry and psychology. There was no representation from non-

white, American or middle class groups, and no intellectual presence from the social sciences or humanities. Spitzer's Task Force was therefore not only culturally and ideologically kindred but also agreed with his broader vision for DSM-III: it should reverse the psychoanalytic tendency (dominating DSM-II) to de-emphasise diagnostic classification in favour of studying the nature and source of intra-psychic conflicts; it should be noncommittal on aetiology, be avowedly descriptive, and should create criteria sets for each disorder in service of securing higher diagnostic reliability.

That the Task Force was unanimous on the aims of DSM-III was clear at its first meeting, as one original member Nancy Andreason later reflected: "When all members of the task force had finished speaking [i.e. setting out their aims], they were clearly astonished at the extent to which they agreed with one another" (Decker 2013: 108). Of course such agreement was not coincidental, but a product of Spitzer's design. Such design was particularly important given the procedures the Task Force adopted to decide the manual's multi-axial structure, criteria sets, diagnostic definitions as well as what disorders to include and/or remove. As neurobiological research played almost no role in guiding how the Task Force settled the aforementioned matters, other procedures were adopted to facilitate the conclusions the Task Force reached. In the following statement, for instance, Spitzer clarifies during interview the procedural grounds upon which 80 disorders were newly added to DSM-III:

[As] psychiatry is unable to depend on biological markers to justify including disorders in the DSM, we looked for other things - behavioural, psychological - we had other procedures....Our general principle was that if

a large enough number of clinicians felt that a diagnostic concept was important in their work then we were likely to add it as a new category. That was essentially it. It became a question of how much *consensus* there was to recognize and include a particular disorder (Interview with author 2012).

It is important to note that the word “consensus” does not necessarily denote the consensus of the wider psychiatric or mental health community. While it was certainly true that wider consensus was tested and sought among psychiatrists (i.e. through questionnaires petitioning selected APA members for views on certain Task Force proposals), the consensus that really mattered was that reached by the Task Force itself, whose prime occupation was to be the final arbiter of any proposed change or inclusion. Task Force consensus could therefore overrule that attained by any advisory committee, lobbying faction or group, a privilege extensively exercised since the demands of different groups would often be at variance, and since the research guiding Task Force decisions was on most matters inconclusive, minimal and/or contradictory, as we will see. Nonetheless, of the research that did exist, its species can be ascertained from archival material held at the APA and from interviews with Task Force members. Such research fell into three categories:

- Questionnaire data gathered via consultation with APA members.
- Existing theoretical or research papers that defined or discussed particular disorders and/or made claims for their wider acceptance.
- Field trials testing the reliability of proposed disorders.

As the problems with the field trials have been extensively studied elsewhere (Kutchins and Kirk 1986, 1994), here we will focus on the Task Force's consultation of questionnaire data and research or theoretical papers. That is, we will focus upon those species of research that helped establish disorders for which reliability was later tested.

The Evidence-Base

To provide an example of the nature of such research, it is instructive to first consider the case of Self-Defeating Personality Disorder (SDPD), which was proposed for inclusion in DSM-III-R and which was ultimately defined as characterizing a pervasive pattern of self-defeating behavior, beginning in early adulthood, and leading the person to avoid pleasurable experiences and seek out situations or relationships in which they would suffer (DSM 1980: 371). This case is important because the controversy surrounding SDPD's proposed inclusion led to its evidence-base being more thoroughly scrutinized than that for almost any other disorder. Extant archival material shows that spirited discussions between the Task Force and consulted critics of SDPD continued until its final inclusion in the appendix of DSM-III-R. Among the critics of SDPD was the Harvard-based psychologist, Paula Caplan, who advised the Task Force on its proposed inclusion but who argued that some of the self-defeating traits proposed as characterizing SDPD were also said to be typical of women who were victims of violence. The diagnosis, it was argued, was therefore dangerous as it could potentially pathologise such women and, at worst, be used in courts of law to suggest that female victims of violence were in fact inviting abuse upon themselves (because they were pathologically disposed to seek out abusive

relationships). This in turn could lead to perpetrators of violence being acquitted on the grounds they were simply complying with their victims' requests.

In the minutes for an undated meeting we hear Task Force members discussing what to make of these criticisms. The Task Force's conclusion is that while the critics were correct to point out that the inclusion of SDPD was a politically fraught issue, they were wrong to suggest the diagnosis might wrongly stigmatise female victims of abuse:

Robert Spitzer: They (the women) present a narrow-gauged but persuasive argument. Their powerful argument is that it is a political hot potato. The feminist issue is a false issue [i.e. that this diagnosis could pathologise female victims of violence].

Fink: Women's arguments seem irrelevant to questions on the table. Are they obscuring their own good arguments? [The 'good arguments' being that SDPD is a controversial diagnosis; the 'irrelevant arguments' being those above posed by Caplan].

Benedek - No empirical basis for category. But you're right – arguments aren't responsive to questions.

Cynthia Rose - we do great disservice by backing off and not acknowledging that this pattern is pathological (Minutes of meeting, undated, Archives, DSM Coll. American Psychiatric Association, Washington D. C.).

Before electing to include SDPD in DSM-III-R, the decision was made to undertake further research into SDPD. Once such research was conducted, it was later scrutinised by Caplan, who discovered it only comprised two studies. The first was conducted by Spitzer himself, and entailed a group of psychiatrists at only one university who already accepted SDPD existed, being shown some old case studies (Caplan 1995: 102). All unanimously agreed the patients in them had SDPD. Caplan argued that just because some psychiatrists at one hospital diagnosed their patients with SDPD was not proof that the disorder actually exists. “All Spitzer’s research proves is that a group of psychiatrists working in the same institution gave the same label – rightly or wrongly - to a given set of behaviours” (Caplan 1995: 205-6).

The second piece involved sending out a questionnaire to a selected number of members of the APA asking them whether the diagnosis SDPD should be included in the DSM. If they voted ‘yes’ then they were asked to describe what they thought the characteristics of SDPD were. If they voted ‘no’ then they were asked to return the questionnaire, blank, without any clinical data. This meant that the only data gathered about the characteristics of SDPD was data obtained from people who believed in the existence of SDPD in the first place. An official report later conducted by the psychologists Kutchins and Kirk showed that only 11 percent of those who returned the questionnaire described what they thought the characteristics of SDPD were (1997). So essentially only 11 percent voted ‘yes’, constituting an unrepresentative sample of the psychiatric community (Kutchins and Kirk, 1997).

As the research basis for SDPD was my most accounts weak, it is important to assess whether this was just an isolated example or whether the evidence-base for other disorders was more robust and extensive. Theodore Millon, a central of the Task Force, addressed this issue when discussing the nature, extent and type of research the Task Force relied upon to reach decisions:

There was very little systematic research, and much of the research that existed was really a hodgepodge—scattered, inconsistent, and ambiguous. I think the majority of us recognized that the amount of good, solid science upon which we were making our decisions was pretty modest (Angell 2009: 29).

Spitzer echoed this concern after my reading Millon's statement during interview:

Well its true that for many of the disorders that were added, there wasn't a tremendous amount of research, and certainly there wasn't research on the particular way that we defined these disorders. In the case of Millon's quote, I think he is mainly referring to the personality disorders.....But again, it is certainly true that the amount of research validating data on most psychiatric disorders is very limited indeed (Interview with author, 2012).

Spitzer then re-emphasised how there was little research not only supporting the inclusion of new disorders, but also supporting how these disorders should be defined:

There are very few disorders whose definition was a result of specific research data. For borderline personality disorder there was some research that looked at

different ways of defining the disorder. And we chose the definition that seemed to be the most valid. But for the other categories rarely could you say that there was research literature supporting the definition's validity (Interview with author, 2012).

Given that Spitzer identifies borderline personality disorder (BDP) as one of the few disorders whose definition was informed by 'some research', it would be useful to inspect what that research comprised. Firstly, while Spitzer's understanding of BDP was highly influenced by clinicians such as Otto Kernberg, Task Force deliberations about how to define and set criteria for BPD transpired periodically for two years without any clear resolution being reached (Decker 2013: 195-202). Henry Pinsker, a member of the original Task Force and the Personality Disorders Work Group, recalled in interview how this impasse was ultimately overcome:

I can remember an early discussion about Borderline, and Spitzer's response was something like 'I just don't know what it is' - but it was a diagnosis that doctors were making. It was only when John Gunderson came up with a paper listing specific characteristics of what he called Borderline Personality Disorder, that we went 'ah ha – we can put this in book'. So the Borderline that went into the DSM was based on Gunderson's paper written sometime in the 1970s (Interview with author, 2013).

When we consult John G. Gunderson's paper today we find it documents what he and his co-author believed to be the six defining features of BPD given their interpretation of the existing literature (Gunderson and Singer, 1975). While their review is

comprehensive in its scope, it does not transcend the usual methodological problems of review-based research, which is susceptible to interpretation bias and unable to control for the multiple variables affecting each study reviewed. Being aware of these methodological difficulties Gunderson later stated: “It had been my expectation at the time this review was completed that it would largely serve as a spring-board for future *empirical* studies which I intended to do” (Gunderson 1983). What Gunderson had not anticipated was the seminal role it would play in defining – via DSM - how BPD would be diagnosed and researched for the next 20 years.

While Gunderson’s article significantly shaped the DSM’s definition of BPD, it was not adopted without amendment. Further changes were discussed and implemented by the Task Force in consultation with Gunderson himself. Since the archival material on these discussions is largely absent from the archives, we are forced to gain a general sense of how such discussions unfolded through the recollections of members. In what follows, Henry Pinsker elaborated during interview on the processes governing not only how the definition BPD was settled, but also on how the definitions of all disorders were reached during Task Force meetings:

We simply discussed things until we were comfortable with it, based on what we as individual members of the group understood and knew. It was really quite primitive compared to what they do now – you know, you read about this DSM-5 Task Force with everything organized around data and studies, whether it be statistical studies or genetic studies, and reconciling this data with that data. We didn’t know about this. It was all about what we knew from clinical experience and from reading. That was it. It was done by consensus of experts, which would

now be considered a very trivial approach, but that's the best we could do then (Interview with author, 2013).

While critics would dispute the levels of robustness Pinsker ascribes to the construction of subsequent DSM editions, when we place his comments alongside those of Spitzer, Millon and Caplan the picture emerges of a Task Force struggling to substantiate its decisions on the basis of solid research. Donald Klein, another member of the original Task Force, also confirmed this struggle in interview:

We had very little in the way of data, so we were forced to rely on clinical consensus, which, admittedly, is a very poor way to do things. But it was better than anything else we had (Interview with author, 2012).

Klein then described how such meetings typically unfolded when consensus was sought in the absence of guiding research:

We thrashed it out, basically. We had a three-hour argument. There would be about twelve people sitting down at the table, usually there was a chairperson and there was somebody taking notes. And at the end of each meeting there would be a distribution of events. And at the next meeting some would agree with the inclusion, and the others would continue arguing. If people were still divided, the matter would be eventually decided by a vote...that is how it went (Interview with author, 2012).

Henry Pinsker also confirms the Task Force's reliance on voting, especially when disagreements ensued:

I don't have specific recollections, some things were discussed over a number of different meetings, [which would sometimes be] followed by an exchange of memoranda about it, and then there would simply be a vote... people would raise hands, there weren't that many people (Interview with author, 2012).

Regarding the legitimacy of this method Pinsker continued: "We never had any question that that is how we would proceed. I had no reservations at all about working that way". On such taskforce meetings, according to Pinsker, there was "certainly more consensus than there was closely contested votes". He continued: "if there was considerable discussion the outcome would be: yeh okay, that makes sense let's do it that way. I think it was more that kind of discussion, than, alright, let's stop discussing it and put it to a vote" (Interview with author, 2013).

The documentary evidence somewhat contradicts Pinsker's claims about how regularly votes were cast, because it shows that Pinsker did not attend all Task Force meetings and so cannot be considered a reliable source on the frequency of votes. Furthermore, Pinsker was not privy too all the voting conducted on conference calls (Decker 2013: 195), and later through the use of ballot letters. Finally, from inspecting the 12 extant minuted Task Force meetings in the APA archives, we find that voting took place in 10 of them. These votes were undertaken on all manner of proposals. For instance, the minutes of votes conducted on the Task Force on 4th Oct 1978 include statements such as:

The taskforce voted to approve a statement drafted after a meeting with the committee on Confidentiality, stating that the first three axes of DSM-III will constitute the official diagnosis, and axes VI and V may be used in ‘special clinical, education and teaching settings’ (Minutes of Task Force Meeting Oct 4th 1978, Archives, DSM Coll. American Psychiatric Association, Washington D. C.).

In the same meeting the Task Force also voted to change the name Introverted Personality Disorder to Schizoid Personality Disorder; to sustain the names of Borderline Personality Disorder, Compulsive Personality Disorder and Passive Aggressive Personality Disorder; they also voted to leave Emancipation Disorder ‘in for the time being’; to name Chronic Affective Disorders as Chronic Hypomanic Disorder, Chronic Depressive Disorder and Cyclothymia. They also voted to expand the definition of Cyclothymic Disorder to include swings of mood of short (hours) duration and voted to include Hysteriod Dysphoria under Atypical Affective Disorders. Finally, as the minutes state: ‘In the course of a rather heated discussion it became clear that (there were questions regarding) the placement and content of Schizoaffective Disorder. In a straw vote the Task Force indicated the desirability of maintaining a separate code for Schizoaffective Disorder’ (Minutes of Task Force Meeting Oct 4th 1978, Archives, DSM Coll. American Psychiatric Association, Washington D. C.).

It is regrettable that today we possess only the minuted conversations that preceded each vote rather than recorded ones, as it is difficult to ascertain from minutes the

precise time spent in deliberating any proposal before a vote was cast. Nonetheless, the minutes do suggest that such discussions were often of only short duration. For instance, in an undated minuted document we find a typical example of how pre-voting discussions unfolded. Here the Task Force was trying to decide the diagnostic criteria for the diagnosis of Substance Dependence:

How many items do you need for a diagnosis of (substance) dependence?

Robbins – persistent cognitive or physical impairment after stopping substance use.

Senay – items for amount of substance consumed.

Maybe in text say something about how some subjects can't give the information, but if they have some common complication such as cirrhosis, one can assume that meets other criteria.

Jaffe – other evidence for other drug abuse, such as needle tracks enable you to make a diagnosis without meeting criteria....perforated nasal septum, Korsakoffs, etc.

Signs and symptoms that are presumptive evidence of use in the past.

Solution: 3 out of 10 symptoms, or any of (a list of medical complications that are definitive evidence of excessive use in the past) ((Minutes of Task

Force Meeting, undated, Archives, DSM Coll. American Psychiatric Association, Washington D. C.).

Rarely do the minutes indicate on what grounds any individual participant lobbied for a particular inclusion or position. The minutes mostly document the lobbying statements, not the corroborating research upon which they were supposedly based. While this does not mean that corroborating research was never discussed on the Task Force, it does raise the question about the extent to which opinions were advanced (and indeed realized) without the evidence-base being scrutinized during meetings. That Task Force discussions could regularly stray from the evidence-base, leading to outcomes that were arbitrary, was concerning to at least one leading member of Task Force itself. In a previously unpublished letter written by Donald Klein to Robert Spitzer, Klein expresses serious concern about how Spitzer, unconstrained by decisive evidence, was allowing himself to be too easily guided by personal preference:

I think that this revision (of DSM-III) in general suffers from problems in deciding the overall criteria for accepting or rejecting a [diagnostic] category.

At times you take the stand that if a large group of respected clinicians agree that something exists descriptively then they should be deferred to [i.e. included]....At other times you take the stand that if something has not shown itself to be demonstrably valuable then it should go out [i.e. be removed].... If this criterion were evenhandedly applied, DSM III would shrink substantially [i.e. because many disorders have not shown themselves to be demonstrably valuable]

And yet at other times you seem to feel that categories that actually have the support of a large group of clinicians...should not be accepted, because somehow or other they just don't fit...

Without agreement on these overarching issues, it is not surprising that there's a lot of disagreement with regard to particular categories (Letter from Donald F. Klein to Robert Spitzer, March 19th 1986. Archives, DSM Coll, American Psychiatric Association, Washington D. C.). ¹

One of the mechanisms by which Spitzer managed such disagreements was consulting relevant leaders in the field. Indeed, the APA archives contain many hundreds of letters evidencing the huge amount of time and energy expended in such communications. These letters contain views being traded back and forth between Spitzer and colleagues on matters as diverse as developing diagnostic definitions, refining diagnostic criteria, and distinguishing separate categories. Not all such correspondence led to concrete inclusions or even amendments, as it was at the discretion of Spitzer and to a lesser extent his Task Force as to whether any given suggestion would be accepted for vote. However, when Spitzer was particularly impressed by the views of a correspondent that person would sometimes be asked to either take up advisory role or to meet Spitzer for a face-to-face meeting. As to why Spitzer would approve some suggestions rather than others, can now only be considered on a case-by-case basis given what we know about his clinical and theoretical beliefs and what extant archival and interview data is able to reveal. What we can ascertain with greater accuracy, however, is how some such face-to-face meetings sometimes unfolded. In an article published by Alix Spiegel, for example,

we gain important insight into how two new disorders ('factitious disorder' and 'brief reactive psychosis'), entered DSM-III through such consultations. We also learn how freely Robert Spitzer was prepared to act alone, creating singlehandedly diagnostic criteria and definitions later put to Task Force vote. The excerpt is worth quoting in full:

Roger Peele and Paul Luisada, psychiatrists at St. Elizabeths Hospital, in Washington, D.C., wrote a paper in which they used the term "hysterical psychoses" to describe the behavior of two kinds of patients they had observed: those who suffered from extremely short episodes of delusion and hallucination after a major traumatic event, and those who felt compelled to show up in an emergency room even though they had no genuine physical or psychological problems. Spitzer read the paper and asked Peele and Luisada if he could come to Washington to meet them. During a forty-minute conversation, the three decided that "hysterical psychoses" should really be divided into two disorders. Short episodes of delusion and hallucination would be labelled "brief reactive psychosis," and the tendency to show up in an emergency room without authentic cause would be called "factitious disorder." "Then Bob asked for a typewriter," Peele says. To Peele's surprise, Spitzer drafted the definitions on the spot. "He banged out criteria sets for factitious disorder and for brief reactive psychosis, and it struck me that this was a productive fellow! He comes in to talk about an issue and walks away with diagnostic criteria for two different mental disorders!" Both factitious disorder and brief reactive psychosis were included in the DSM-III with only minor adjustments (Spiegel 2005).

DSM-III was published in 1980, and while its global influence on mental health research and practice had been assured by the end of the 1980s, the central mechanisms by which it was established remained largely unknown to most professionals and the public alike. As the decade unfolded considerable uncertainty grew concerning the extent to which biological evidence or solid research supported the definitions of the disorders contrived, the validity of the disorders included, and the symptom thresholds people must meet to receive the diagnosis. Most professionals did not know that such matters were not established upon a clear evidence-base, but were the product of committee consensus, which, at best, reflected the well-meaning and culturally-embedded professional opinions of a small subset of psychiatrists. Indeed, as Spitzer stated in interview: ‘Our team was certainly not typical of the psychiatry community, and that was one of the major arguments against DSM III: it allowed a small group with a particular viewpoint to take over psychiatry and change it in a fundamental way’. When asking Spitzer what he made of that criticism, he responded: “What did I think of that charge? - Well, it was absolutely true. It was a revolution, that’s what it was. We took over because we had the power” (interview with author, 2012).

The prevalence of such methodological problems sits uneasily with the manner in which the APA still represents the construction of DSM-III. Statements made by the APA today still stress, for example, that DSM-III’s construction was “facilitated by extensive empirical work on the construction and validation of explicit diagnostic criteria” (APA 2014), and elsewhere that “many revisions appearing in DSM-III-R were based on over 2,000 scientific publications which cited DSM-III” (Halleck et al.,

1992). The suggestion that DSM-III was rigorously evidence-based jars with both the interview and archival data set down above.

DSM-III's influence on DSM-IV:

In 1994 Spitzer's DSM was replaced by DSM-IV. During two separate interviews in 2012 with DSM-IV's chairperson, Allen Frances, he discussed the extent to which DSM-IV departed from its preceding edition, DSM-III:

DSM IV was a remarkably unambitious and modest effort to stabilize psychiatric diagnosis, and not to create new problems. This meant keeping the introduction of new disorders to an absolute minimum (interview with author, 2012).

Here Frances is referring to the Task Force adding only eight new disorders to the main manual, which is a modest amount considering that Spitzer's Task Force introduced eighty. From another standpoint, however, Frances' claim that 'new disorders were kept to an absolute minimum', ignores that DSM-IV placed an additional 30 new disorders in the appendix for 'further study', and re-coded and subdivided many existing disorders. When including these appendix disorders and subdivisions, all of which patients can be diagnosed with, Frances's Task Force actually expanded the DSM from 292 to 374 disorders.

Where Frances' claim to modesty stands up to further scrutiny is in noticing that the DSM-IV Task Force only significantly reformulated four of the 292 disorders inherited from DSM-III-R. This meant that most of DSM-III's categories were

imported into DSM-IV without significant alteration. When reflecting on why so much of DSM-III was transferred wholesale into DSM-IV, Frances was explicit about the rationale:

If we were going to either add new diagnosis or eliminate existing ones there had to be substantial scientific evidence to support that decision. And there simply wasn't. So by following our own conservative rules we couldn't reduce the system anymore than we could increase it. Now, you could argue that is a questionable approach, but we felt it was important to stabilise the system and not make arbitrary decisions in either direction (interview with author, 2012).

An obvious question regarding this rationale is whether it wrongly assumed that the disorders DSM-III included, their definitions and criteria sets, were themselves established on a robust evidence base? Frances addressed this issue in the following way:

We did not assume that at all. We knew that everything that came before was arbitrary [Frances corrected himself]; we knew that *most* decisions that came before were arbitrary. I had been involved in DSM III. I understood its limitations probably more than most people did. But, the most important value at that time, was to stabilise the system not change it arbitrarily... it felt better to stabilise the existing arbitrary decisions than to create a whole assortment of new ones (interview with author, 2012).

Frances's admissions, coupled with textual comparisons of both manuals, indicate that the influence of DSM-III did not wane after 1994. Its influence lived on in three central ways: by establishing and normalising the procedures of voting and consensus upon which later editions also came to depend; by providing the benchmark against which changes were made by subsequent Task Forces; and by establishing most of the disorder definitions and criteria sets still in use today. While it is beyond the scope of this article to document the many diagnostic reformulations that were conducted over DSM-IV, DSM-IV-R and DSM-5, as well as to discuss the removal the multi-axial system from DSM-5, knowledge of the procedures by which DSM-III and DSM-III-R were developed suggests that critical scrutiny of subsequent manuals should not solely focus on the merits or demerits of newly introduced changes, as has been largely the case with criticisms of DSM-5, but also upon the extent to which deploying and incorporating DSM-III-based procedures and categories was legitimate by any robust empirical standard. In short, important questions still remain about the viability of the DSM's cumulative method, insofar as what has been accumulated over consecutive editions was the product of processes to prone to error.

Conclusion

While it is broadly accepted among most socio-cultural scholars of psychiatry that the DSM is a culturally constructed document (e.g. Gaines 1992; Littlewood 1992; Good 1996; Kleinman 2012), solid empirical evidence concerning the centrality of voting-based consensus has, interestingly, remained elusive. This article has attempted to redress this omission by expanding the evidence-base concerning the centrality and prevalence of voting on DSM Task Forces, thus substantiating that the contrivance of DSM categories was largely facilitated by **invested cultural negotiation, rather**

than by the consultation of what today's research environment would regard as satisfactory empirical evidence. Where this article therefore makes its contribution is by sourcing new oral and archival data illustrating that the separate disorders into which DSM-III organized diverse behavioral and mental phenomena were largely the outcome of vote-based judgments settled by a small, culturally-homogenous subset of mental health professionals who were socially positioned at a given moment in psychiatric history to have their judgments ratified by the institutional apparatus of the APA. While such judgments may indicate that some professionals with similar socio-cultural beliefs and interests see some things in the same way at given point in time, they do not confirm that what they see is either objectively true or stable in any verifiable sense. So while the extensive time and dedication invested in the making of DSM-III may warrant professional respect, the finished product does not warrant uncritical deference insofar as further analysis may reveal more about the professional and broader culture of its particular creators than about the phenomena upon which they pronounced.

References Cited

Angell, M (2009) Drug Companies & Doctors: A Story of Corruption. *The New York Review of Books*, January 15.

APA (2014) DSM: *History of the Manual*. American Psychiatric Association. <http://www.psychiatry.org/practice/dsm/dsm-history-of-the-manual>, accessed June 21st 2014).

Caplan, P (1995) *They Say You're Crazy*. New York: DaCapo Press.

Conrad, P (2007) *The Medicalization of Society: The Transformation of Human Conditions into Treatable Disorders*. Baltimore: Johns Hopkins University Press.

Cooper, J et al. (1972) *Psychiatric Diagnosis in New York and London*. Maudsley Monograph No 20. London: Oxford University Press.

Cooper, R (2004) What is Wrong with the DSM? *History of Psychiatry* 15(1): 005–025.

Davies, J (2013) *Cracked: why psychiatry is doing more harm than good*. London: Icon Books.

Decker, H (2013) *The Making of DSM-III: a Diagnostic Manual's Conquest of American Psychiatry*. New York: Oxford University Press.

Diagnostic and Statistical Manual of Mental Disorders (Third Edition) (1980) Washington: American Psychiatric Association.

Diagnostic and Statistical Manual of Mental Disorders (Third Edition - Revised) (1985) Washington: American Psychiatric Association.

Eysenck, H (1986) A Critique of Contemporary Classification and Diagnosis. In: *Contemporary Directions in Psychopathology: Toward the DSM-IV*. Millon, T., and Klerman, G.L. (eds): 73-98 New York: Guilford.

Gaines, A, D. (1992) From DSM-I to III-R; voices of self, mastery and the other: a cultural constructivist reading of U.S. psychiatric classification *Soc Sci Med* 35(1):3-24.

Good, Byron (1996) Culture and DSM-IV: Diagnosis, knowledge and power. *Culture, Medicine and Psychiatry* 20 (2): 127 – 132

Gunderson, J. G. and Singer M. T. (1975) Defining Borderline Personality Disorder: an Overview. *The American Journal of Psychiatry* 132: 1-10.

Gunderson, J. G. (1983) Untitled Interivew. *Current Contents* 19:20

Gurland, B. J. et al. (1972) On the use of Psychiatric Diagnosis for Comparing Psychiatry Populations. *Psychiatry Q* 46: 461-473.

Halleck, S. L. et al (1992) *Use of Psychiatric Diagnosis in the Legal Process*. Task Force Report 32, American Psychiatric Association, 1992.

Junod, S. W. (2008) FDA and Clinical Drug Trials: A Short History. *In A Quick Guide to Clinical Trials*. Madhu Davies and Faiz Kerimani, eds. Pp. 25-55 Washington: Bioplan Inc.

Kirk, S. A. and H. K. (1992) *The Selling of DSM: The Rhetoric of Science in Psychiatry*. New York: Aldine de Gruyter.

Kleinman Arthur (2012) Rebalancing Academic Psychiatry: why it needs to happen – and soon. *The British Journal of Psychiatry* 201: 421-422

Kutchins, H. and Stuart A. K. (1986) The Reliability of DSM-III: A Critical Review. *Social Work Research and Abstracts* 22: 3-12.

Kutchins, H. and Stuart A. K. (1988) The Business of Diagnosis: DSM-III and Clinical Social Work. *Social Work* 33(3): 215-220.

Kutchins, H. and Stuart A. K. (1994) The Myth of Reliability. *Journal of Mind and Behavior* 15 (1 and 2): 71-86.

Kutchins, H. and Stuart A. K. (1997) *Making us Crazy*. New York: Free Press.

Lane, C. (2007) *Shyness: How Normal Behavior Became a Sickness*. New Haven: Yale University Press.

Littlewood , R. (1992) DSM-IV and culture: Is the classification internationally valid? *Psychiatric Bulletin* 16 : 257–261.

Rosenhan, David L.

1973 On Being Sane in Insane Places. *Science* 179 (4070): 250-258.

Scheff, T. J. (1986) Accountability in Psychiatric Diagnosis: A proposal. *In Contemporary Directions in Psychopathology: Toward the DSM-IV*. Theodore Millon and Gerald L. Klerman eds: 265-278 New York: Guilford.

Shorter, E. (2009) *Before Prozac: the Troubled History of Mood Disorders in Psychiatry*. Oxford: Oxford University Press.

Spiegel, A. (2005) The Dictionary of Disorder: How One Man Revolutionized Psychiatry. *The New Yorker* January 3rd.

Szasz, T. (2007) *The Medicalization of Everyday Life (Selected Essays)*. Syracuse: Syracuse University Press.

Tietze, T (2014) Neoliberal Psychiatry and its Discontents. *Internaltional Soc*, 141.

Wilson, M. (1993) DSM and the Transformation of American Psychiatry: a history. *American Journal of Psychiatry*, 150(3): 399-410.

¹ Documents such as these, coupled with the interview data gathered by Alex Speigal (2005) and the executive power afforded to Robert Spitzer by the APA to decide the composition of the Task Force and working groups and to determine what subjects came before the Task Force for scrutiny, suggests that DSM III significantly reflected the opinions and preferences of this single individual.